

Санкт-Петербургский государственный университет
Кафедра моделирования экономических систем

Иванов Никита Григорьевич

Выпускная квалификационная работа бакалавра

Анализ горизонта временных рядов

Направление 010400

Прикладная математика и информатика

Научный руководитель,
доктор физ.-мат. наук,
профессор
Прасолов А. В.

Санкт-Петербург
2016

Содержание

Введение	3
Постановка задачи	5
Глава 1. Построение тренда временного ряда	6
1.1. Выделение линейной составляющей	6
1.2. Разложение в ряд Фурье	7
1.3. Оптимизация коэффициентов ряда Фурье	9
Глава 2. Эмпирический анализ горизонта прогнозирования времен- ных рядов	14
2.1. Алгоритм построения горизонта прогнозирования	15
2.1.1. Пример 1. Нефть марки «Brent»	16
2.1.2. Пример 2. Отношение цен евро к доллару	17
2.1.3. Пример 3. Температура в Санкт-Петербурге	18
2.2. Алгоритмы построения среднего горизонта прогнозирования	20
Заключение	23
Список литературы	24
Приложение	26
Программная реализация в пакете MATLAB	26

Введение

Задача прогнозирования временных рядов является необходимым инструментом принятия решений практически во всех сферах нашей жизни. В экономике это курсы валют, цены на нефть, стоимости акций. В медицине характерным примером является кардиограмма. В социологии временные ряды используют для рассмотрения мнения населения в различных вопросах с течением времени. В биологии с помощью временных рядов рассматривают численность популяции на больших промежутках времени. Наконец, в географии рассматривается температура, атмосферное давление, скорость ветра.

Дадим несколько определений для одинакового понимания терминов.

Определение 1. Случайным (стохастическим) процессом будем называть однопараметрическое семейство случайных величин, которые имеют конечную дисперсию.

Определение 2. Реализацией процесса (выборочной функцией, выборкой, траекторией) назовём соответствующее семейство реализаций случайных величин при последовательном переборе всех параметров.

Определение 3. Трендом назовём линию математических ожиданий случайных величин стохастического процесса.

Определение 4. Временным рядом является одна из реализаций случайного процесса.

Определение 5. Термином «прогноз» будем обозначать модель тренда случайного процесса, построенного по временному ряду, и полосы возможных отклонений вокруг тренда (модели). При необходимости полосу отклонений можно или совсем не рассматривать, или принимать во внимание какую-либо её процентную часть.

Определение 6. Горизонтом прогнозирования временного ряда называется интервал в будущем, на который делается прогноз по прошлому.

Определение 7. Глубиной прогноза временного ряда будем называть интервал в прошлом, по которому определяется горизонт прогнозирования.

Относительно приведенных определений имеется подробный учебник Острема К. Ю. [1].

Существуют две основных модели, с помощью которых производится прогнозирование временных рядов: авторегрессионная и модель с выделе-

нием тренда. Проблемой первой модели является то, что в силу своего построения прогнозировать можно только на краткосрочные периоды, т.к. дисперсия для каждого последующего временного узла увеличивается в разы, то есть результат прогнозирования становится всё менее достоверным. Тем временем вторая модель активно используется в долгосрочном прогнозировании. Такие модели рассматривает, например, Т. Андерсон [2].

Зачастую авторы позволяют себе делать прогноз на любые интервалы — краткосрочные и долгосрочные, не задумываясь при этом, как далеко можно прогнозировать. В литературе практически отсутствуют исследования об оценке горизонта прогнозирования временного ряда. Данной проблемой занимались Прасолов А. В., Хованов Н. В и Вей К. [3–5].

Постановка задачи

Продолжая начинания вышеуказанных авторов, в работе будет решаться задача определения оценки сверху горизонта прогнозирования временного ряда. Требуется выделить тренд временного ряда методом разложения в ряд Фурье. Необходимо учесть неточность вычисления коэффициентов ряда Фурье и свести её к минимуму, оптимизировав коэффициенты [6]. Далее стоит задача — на основе выделенного тренда получить алгоритм построения прогноза и горизонта временного ряда [7]. Работу алгоритма следует проверить на временных рядах различной природы: цены на нефть марки «Brent», отношение курсов валют евро к доллару, температурные данные в Москве и Петербурге, а также мировые цены на зерно. Как результат, необходимо сравнить полученные оценки горизонта на соответствие реальной жизни.

Глава 1. Построение тренда временного ряда

Необходимо понимать, что тренд представляет из себя сумму детерминированной (линейной $Y(T_i)$ и Фурье-модели $F(T_i)$) и случайной (спекулятивной) составляющей $\omega(T_i) \neq 0$. В данной главе происходит построение в три шага детерминированной части. Для наглядности результатов все расчёты произведены, используя цены на нефть марки «Brent», взятые в период с января 2009 г. по сентябрь 2014 г. с интервалом одна неделя [8].

Рассмотрим временной ряд $A(T_i), i = \overline{1, n}$, где T_i — равноотстоящие временные замеры. В контексте данного примера $T_i - T_{i-1} = 1, i = \overline{2, n}$; $n = 299$; $A(T_i)$ — цена в каждый момент времени.

1.1. Выделение линейной составляющей

Сначала выделим линейную составляющую тренда с помощью метода наименьших квадратов. Приближим $A(T_i), i = \overline{1, n}$, полиномом первой степени $Y(t) = at + b$.

Для временного ряда, показанного на рис. 1, линейная функция будет иметь вид

$$Y(t) = 0,19t + 68,0.$$

Подставив в $Y(t)$ вместо непрерывной переменной t дискретные значения $T_i, i = \overline{1, n}$, получим (рис. 2)

$$Y(T_i) = 0,19T_i + 68,0, \quad i = \overline{1, n}.$$

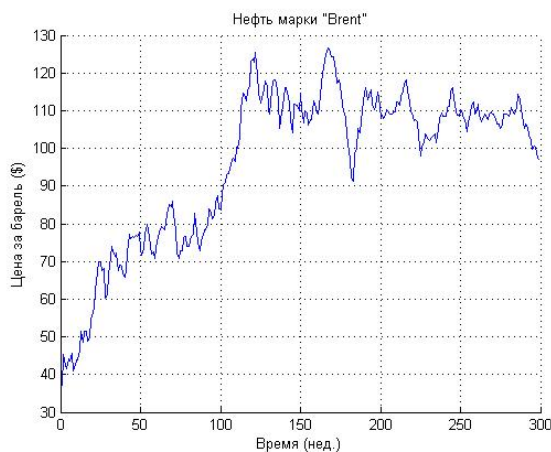


Рис. 1: Цена на нефть марки «Brent»

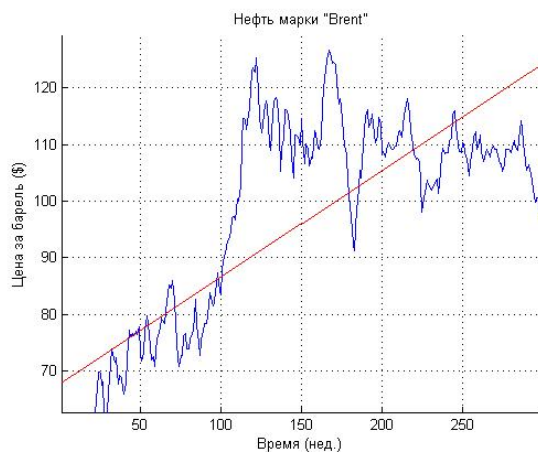


Рис. 2: Линейная составляющая $Y(T_i)$

Прежде чем приступить к следующему шагу, вычтем значения $Y(T_i)$ из $A(T_i)$ и получим новые значения остатков (рис. 3)

$$R(T_i) = A(T_i) - Y(T_i), i = \overline{1, n}.$$

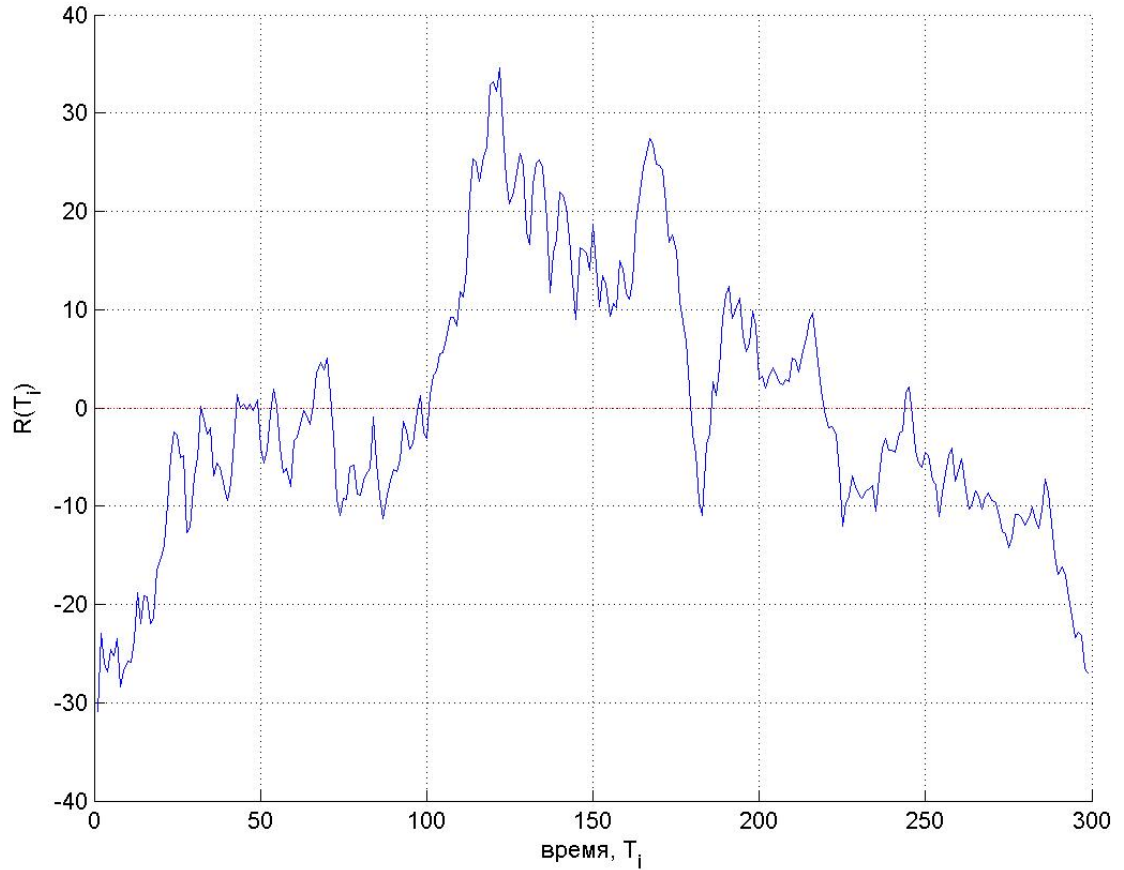


Рис. 3: $R(T_i)$ — Временной ряд без линейной составляющей

1.2. Разложение в ряд Фурье

На втором этапе приблизим новый временной ряд $R(T_i), i = \overline{1, n}$, методом разложения в ряд Фурье.

Тригонометрическим рядом Фурье функции $Z(t) \in L^2([0, T])$ называют функциональный ряд вида:

$$Z(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos \left(\frac{2\pi k}{T} t \right) + b_k \sin \left(\frac{2\pi k}{T} t \right) \right).$$

Так как мы имеем дискретное время $T_i, i = \overline{1, n}$, то коэффициенты

a_k и b_k будут искажаться приближённо, с заменой интегралов

$$a_k = \frac{2}{T} \int_0^T R(t) \cos\left(\frac{2\pi k}{T} t\right) dt; \quad b_k = \frac{2}{T} \int_0^T R(t) \sin\left(\frac{2\pi k}{T} t\right) dt$$

площадями соответствующих прямоугольников [2]:

$$a_k \approx \frac{2}{n} \sum_{i=1}^n R(T_i) \cos\left(\frac{2\pi k}{n} T_i\right); \quad b_k \approx \frac{2}{n} \sum_{i=1}^n R(T_i) \sin\left(\frac{2\pi k}{n} T_i\right).$$

В силу того, что мы численно аппроксимируем функцию, количество слагаемых (гармоник) в ряде Фурье будем назначать сами, введя переменную c . Тогда, учитывая тот факт, что a_0 обратится в 0 (это получено после того, как вычли линейную составляющую), имеем

$$F_c(T_i) = \sum_{k=1}^c \left(a_k \cos\left(\frac{2\pi k}{n} T_i\right) + b_k \sin\left(\frac{2\pi k}{n} T_i\right) \right), \quad i = \overline{1, n}, \quad c \in \mathbb{N}.$$

Приведём для наглядности результат для $c = 3$ (рис. 4) и $c = 15$ (рис. 5).

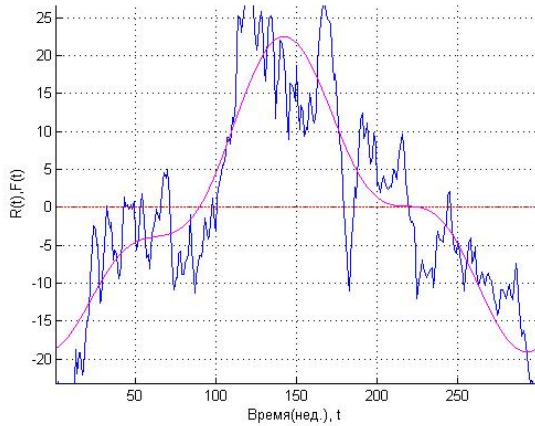


Рис. 4: Периодическая составляющая временного ряда при $c=3$

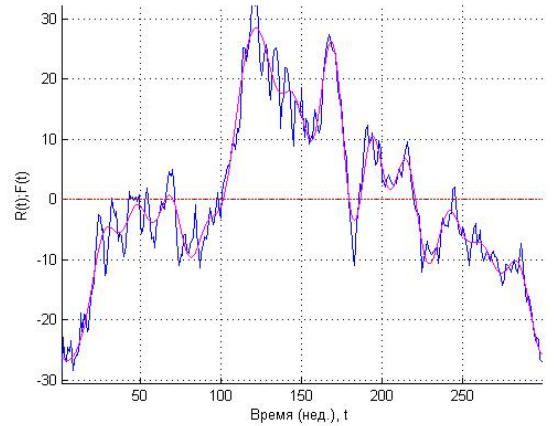


Рис. 5: Периодическая составляющая временного ряда при $c=15$

Рассмотрим квадраты разностей $R(T_i)$ и $F_c(T_i)$ для различных $c = \overline{1, 200}$ (рис. 6)

$$D(c) = \sum_{i=1}^n (R(T_i) - F_c(T_i))^2.$$

Замечание 1. При $c = 149$ функции $F(T_i)$ и $R(T_i)$ фактически совпадают. Притом значения $D(c)$ при $c = 148$ и $c = 150$, $c = 147$ и $c = 151, \dots$ совпадают с точностью до знака. Тогда можно предположить, что оптимальное число гармоник c временного ряда соответствует $N = \frac{n-1}{2}$ для

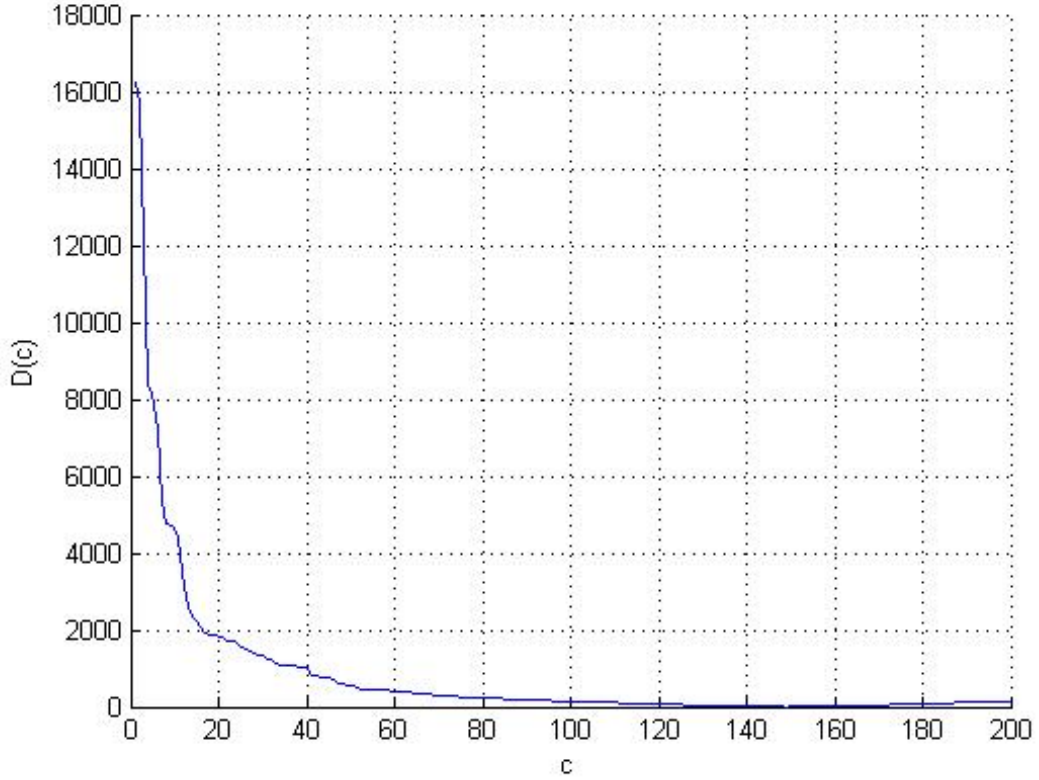


Рис. 6: График $D(c)$

нечётного и $N = \frac{n}{2}$ для чётного количества узлов временного ряда. Такое предположение логично вытекает из того, что время дискретно, значит, нужно приблизить функцию от времени в n точках. Но N — число гармоник ряда Фурье и их достаточно, чтобы приблизить временной ряд с n узлами [2, 10].

1.3. Оптимизация коэффициентов ряда Фурье

В полученной модели $F(T_i)$ присутствует ошибка вычисления, которая возникает из двух источников: приближённое вычисление интегралов по заданной функции $R(T_i)$, которая раскладывается в ряд, и отсутствие этой функции, поскольку под интегралом находятся реализации случайных величин, а не значения детерминированной функции.

Попробуем учесть данный факт в коэффициентах $a_k + \varepsilon_k$, $b_k + \delta_k$ нового ряда Фурье:

$$\sum_{k=1}^N \left((a_k + \varepsilon_k) \cos \left(\frac{2\pi k}{n} T_i \right) + (b_k + \delta_k) \sin \left(\frac{2\pi k}{n} T_i \right) \right), \quad i = \overline{1, n}.$$

Введём функцию $S(\varepsilon_k, \delta_k)$:

$$S(\varepsilon_k, \delta_k) = \sum_{i=1}^n \left(R(T_i) - \sum_{k=1}^N \left((a_k + \varepsilon_k) \cos \left(\frac{2\pi k}{n} T_i \right) + (b_k + \delta_k) \sin \left(\frac{2\pi k}{n} T_i \right) \right) \right)^2.$$

Минимизируем её по неизвестным параметрам ε_k и δ_k . Функция S представляет собой сумму квадратов, значит глобальный минимум этой функции равен нулю, то есть $\min_{\varepsilon_k, \delta_k} S(\varepsilon_k, \delta_k) = 0$. Преобразуем разность, сгруппировав известные значения в одной скобке, а неизвестные со своими известными коэффициентами в другой:

$$S(\varepsilon_k, \delta_k) = \sum_{i=1}^n \left(\left[R(T_i) - \sum_{k=1}^N \left(a_k \cos \left(\frac{2\pi k}{n} T_i \right) + b_k \sin \left(\frac{2\pi k}{n} T_i \right) \right) \right] - \left[\sum_{k=1}^N \left(\varepsilon_k \cos \left(\frac{2\pi k}{n} T_i \right) + \delta_k \sin \left(\frac{2\pi k}{n} T_i \right) \right) \right] \right)^2.$$

Но тогда из того, что $\min_{\varepsilon_k, \delta_k} S(\varepsilon_k, \delta_k) = 0$, следует, что каждое слагаемое под знаком внешней частичной суммы должно быть равно нулю. А из того, что каждое из слагаемых возведено в квадрат, вытекает

$$\begin{aligned} R(T_i) - \sum_{k=1}^N \left(a_k \cos \left(\frac{2\pi k}{n} T_i \right) + b_k \sin \left(\frac{2\pi k}{n} T_i \right) \right) = \\ = \sum_{k=1}^N \left(\varepsilon_k \cos \left(\frac{2\pi k}{n} T_i \right) + \delta_k \sin \left(\frac{2\pi k}{n} T_i \right) \right). \end{aligned}$$

Обозначим

$$\begin{aligned} X(T_i) = R(T_i) - \sum_{k=1}^N \left(a_k \cos \left(\frac{2\pi k}{n} T_i \right) + \right. \\ \left. + b_k \sin \left(\frac{2\pi k}{n} T_i \right) \right), \quad i = \overline{1, n}. \end{aligned}$$

Тогда, подставив все известные значения, получим систему линейных уравнений с $(n \times 2N)$ -матрицей, которую разрешим с помощью метода

наименьших квадратов.

$$\begin{pmatrix} \cos\left(\frac{2\pi 1}{n}T_1\right) & \sin\left(\frac{2\pi 1}{n}T_1\right) & \cdots & \cos\left(\frac{2\pi N}{n}T_1\right) & \sin\left(\frac{2\pi N}{n}T_1\right) \\ \cos\left(\frac{2\pi 1}{n}T_2\right) & \sin\left(\frac{2\pi 1}{n}T_2\right) & \cdots & \cos\left(\frac{2\pi N}{n}T_2\right) & \sin\left(\frac{2\pi N}{n}T_2\right) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cos\left(\frac{2\pi 1}{n}T_n\right) & \sin\left(\frac{2\pi 1}{n}T_n\right) & \cdots & \cos\left(\frac{2\pi N}{n}T_n\right) & \sin\left(\frac{2\pi N}{n}T_n\right) \end{pmatrix} * \begin{pmatrix} \varepsilon_1 \\ \delta_1 \\ \vdots \\ \varepsilon_N \\ \delta_N \end{pmatrix} = \begin{pmatrix} X(T_1) \\ X(T_2) \\ \vdots \\ X(T_n) \end{pmatrix}$$

Обозначим за Q матрицу данной системы, y — вектор искомых коэффициентов, X — вектор-столбец значений. В данных обозначениях система примет вид

$$Qy = X$$

Домножим систему на Q^T слева: $Q^T Qy = Q^T X$ и, уже имея квадратную невырожденную матрицу $Q^T Q$ размерности $(2N \times 2N)$, находим вектор y .

Итак, получены коэффициенты $\varepsilon_1, \delta_1, \dots, \varepsilon_N, \delta_N$. Обозначим

$$\varepsilon(T_i) = \sum_{k=1}^N \left(\varepsilon_k \cos\left(\frac{2\pi k}{n}T_i\right) + \delta_k \sin\left(\frac{2\pi k}{n}T_i\right) \right), \quad i = \overline{1, n}.$$

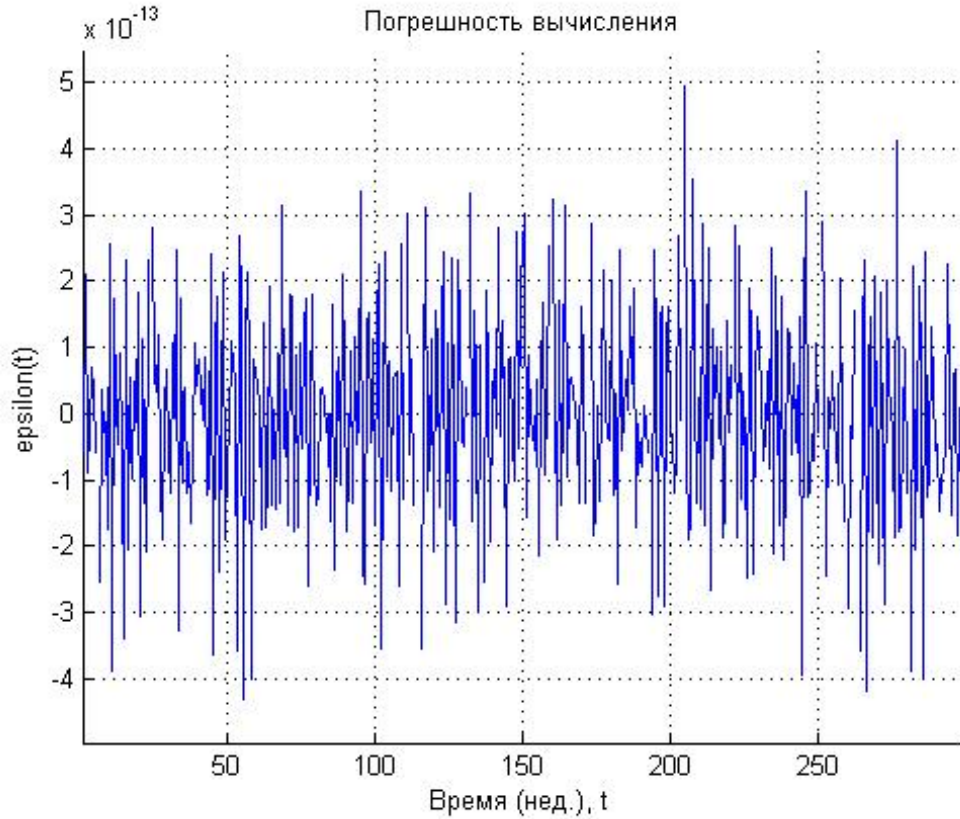


Рис. 7: Погрешность $\varepsilon(T_i)$

Из графика (рис. 7) видно, что введение новых коэффициентов при $c = N$ практически не повлияло на аппроксимацию (ошибка погрешности вычисления коэффициентов не превышает $5 * 10^{-13}$), что и логично, ведь $D(N) \approx 0$. Однако, как известно, временные ряды используются для прогнозирования, где требуется составить прогноз на основе тренда, а не интерполяции. Значит, такая хорошая точность на известном интервале может быть излишней. Поэтому в задаче прогнозирования разумно выбирать $c < N$. Рассмотрим функции $\tilde{X}_c(T_i)$ и $\tilde{\varepsilon}_c(T_i)$:

$$\begin{aligned}\tilde{X}_c(T_i) &= R(T_i) - \sum_{k=1}^c \left(a_k \cos \left(\frac{2\pi k}{n} T_i \right) + b_k \sin \left(\frac{2\pi k}{n} T_i \right) \right) = \\ &= R(T_i) - F_c(T_i), \quad i = \overline{1, n}.\end{aligned}$$

$$\tilde{\varepsilon}_c(T_i) = \sum_{k=1}^c \left(\varepsilon_k \cos \left(\frac{2\pi k}{n} T_i \right) + \delta_k \sin \left(\frac{2\pi k}{n} T_i \right) \right), \quad i = \overline{1, n}.$$

Тогда введём функцию

$$E(c) = \sum_{i=1}^n \left(\tilde{X}_c(T_i) - \tilde{\varepsilon}_c(T_i) \right)^2.$$

Из графика (рис. 8) видно, что $E(c) < D(c)$, $c = \overline{1, N}$.

Это значит, что выделение в коэффициентах Фурье случайной составляющей целесообразно и обеспечивает лучшую сходимость.

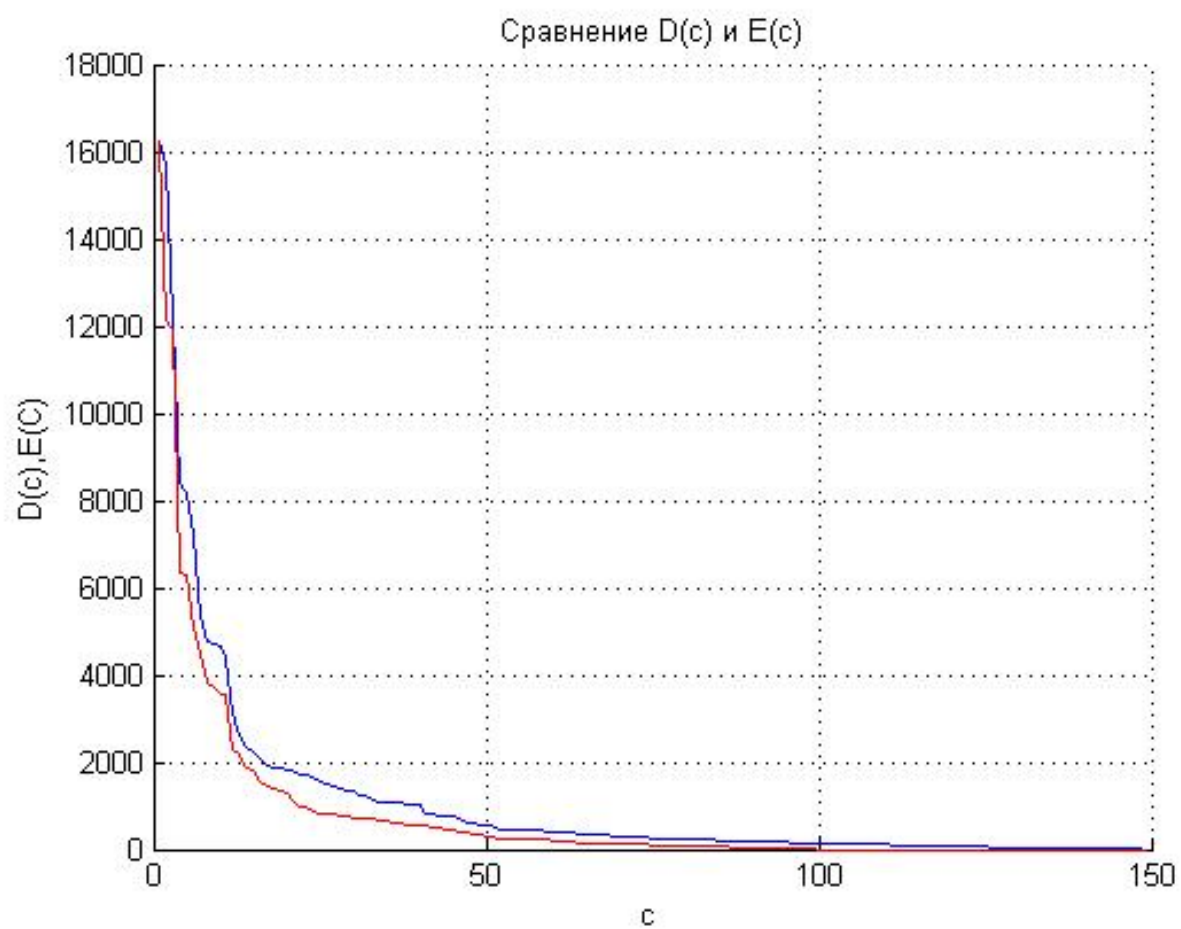


Рис. 8: Графики $D(c)$ и $E(c)$

Глава 2. Эмпирический анализ горизонта прогнозирования временных рядов

Целью настоящей главы является оценка сверху горизонта прогнозирования, т.е. интервала в будущем, на который делается прогноз по прошлому. Данные для эмпирического анализа взяты из открытых источников по экономике и метеорологии [8, 11, 12]. Попутно будет решаться задача окончательного определения детерминированной части тренда временного ряда, а именно определения оптимального числа гармоник с функции $F_c(T_i)$. Переопределим в этой главе $F_c(T_i)$ как сумму линейной составляющей и ряда Фурье с оптимизированными коэффициентами

$$F_c(T_i) = Y(T_i) + \sum_{k=1}^c \left((a_k + \varepsilon_k) \cos \left(\frac{2\pi k}{n} T_i \right) + (b_k + \delta_k) \sin \left(\frac{2\pi k}{n} T_i \right) \right), \quad i = \overline{1, n},$$

Ряды Фурье как инструмент аппроксимации функции довольно быстро будут сходиться к функции, построенной на выборке. Целью моделирования является построение аппроксимирующей функции, стремящейся к тренду, а не к выборке. Однако есть возможность вокруг Фурье-модели построить полосу, которая будет оценкой расположения тренда. Для этого предлагается построить такую часть ряда Фурье, полоса остатков относительно которой будет не меньше, чем полоса первых разностей исходного временного ряда. Ниже это будет показано более подробно.

Однако $F_c(T_i)$ является лишь детерминированной составляющей модели, и нельзя забывать, что существует случайная составляющая $\omega(T_i) \neq 0$. Таким образом, на данный момент приближена рядом Фурье выборка $A(T_i)$, но не тренд. Задача — по выборке стохастического процесса сделать прогноз на какой-то интервал вперёд, используя эмпирические данные.

Проделан анализ временных рядов различной природы: данные цен на нефть марки «Brent», отношение курсов валют евро к доллару и ежедневные температурные данные в Санкт-Петербурге. У одних видов рядов может быть периодичность, у других нет. Временные ряды, отражающие разные процессы в экономике и, более широко, в природе, характеризуются разной степенью «консерватизма»: стационарные стохастические процес-

сы имеют независимое от времени распределение случайных величин, это самый простой для прогнозирования случай. Более сложны для прогнозирования процессы с переменным математическим ожиданием, что требует моделирования тренда. В данной работе используется моделирование с помощью разложений в ряд Фурье, предполагая при этом, что случайные величины процесса имеют ограниченную область определения, и, кроме того, что тренд непрерывен и липшицев.

2.1. Алгоритм построения горизонта прогнозирования

Рассмотрим один из самых неудобных для прогнозирования экономический параметр — цены на нефть марки «Brent», взятые в период с января 2009 г. по сентябрь 2014 г. с интервалом одна неделя [8]. Данный временной ряд содержит в себе экономические характеристики сложной природы: экономическую составляющую, которая является следствием равновесия между спросом и предложением нефти, а так же спекулятивную составляющую, которая подвержена огромному количеству различных возмущений. Следствием этого явилось то, что прогнозирование данных нефти затруднительно.

Продолжим рассматривать временной ряд для нефти марки «Brent» $A(T_i)$. Делаем предположение о том, что для каждого двух рядом стоящих реализаций случайной величины временного ряда $A(T_i)$ дисперсии являются конечными и притом не сильно различающимися, то есть тренд меняется «плавно». Тогда, как предлагали Бокс Дж. и Дженкинс Г. [9], построим временной ряд первых разностей $D(T_i)$ на $2/3$ данных (рис. 9): $n_1 = 2/3n$; $D(T_i) = A(T_{i+1}) - A(T_i)$, $i = \overline{1, n_1 - 1}$.

На графике «звёздочками» отмечены максимальное и минимальное значения ряда. Половина размаха ряда $D(T_i)$ и будет являться шириной полосы, которая строится вокруг тренда. Следовательно, возникает задача построения тренда. При построении $F(T_i)$ необходимо выбрать количество гармоник s .

Замечание 2. В силу своего построения узлы временного ряда $D(T_i)$ являются реализацией случайной величины с дисперсией в два раза больше, чем у ряда $A(T_i)$. Для того, чтобы использовать размах ряда $D(T_i)$ в дальнейшем, необходимо привести дисперсию к тем же размерам, что и у

ряда $A(T_i)$, поэтому за ширину полосы вокруг тренда выбирается половина размаха ряда $D(T_i)$.

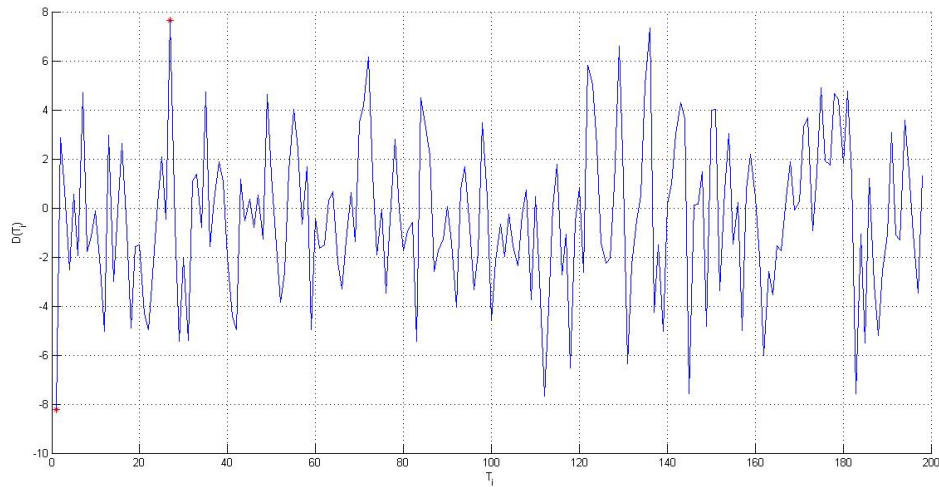


Рис. 9: $D(T_i)$ для нефти марки «Brent»

Алгоритм построения горизонта:

1. Строим на n_1 данных $F(T_i)$ по одной гармонике.
2. Рассматриваем ряд остатков $X(T_i) = A(T_i) - F(T_i)$ на n_1 данных. Из этого ряда нам понадобится размах.
3. Сравниваем его с размахом ряда $D(T_i)$: если половина размаха ряда $D(T_i)$ меньше размаха ряда $X(T_i)$, то увеличиваем число гармоник на единицу и возвращаемся к п.1., иначе останавливаемся и строим модельный прогноз.
4. Строим полосу ширины половины размаха ряда $D(T_i)$ вокруг модельных данных $F(T_i)$ как на идентификационной, так и на прогнозной части ряда T_i .
5. Определяем выход контрольных данных из полосы, окружающей модельные данные на последней трети ряда T_i . Этот выход и будет являться горизонтом прогнозирования.

Как и в предыдущей главе, для наглядности покажем на трёх примерах алгоритм построения оценки прогноза.

2.1.1. Пример 1. Нефть марки «Brent»

В результате использования алгоритма получено, что наименьшее необходимое количество гармоник — 29. При этом горизонт прогноза оказался равен 1 (рис. 10). Линейная составляющая функции $F(T_i)$ имеет вид:

$Y(T_i) = 0,36793 T_i + 53,145$. Такой небольшой горизонт говорит о том, что прогнозирование цен на нефть на большие периоды данным методом даст неверный результат.

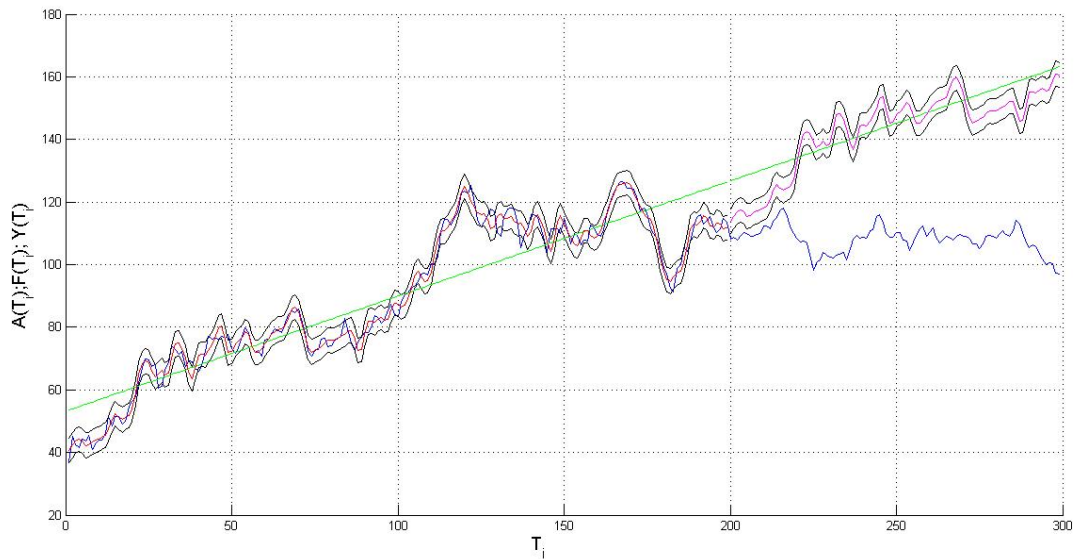


Рис. 10: $F(T_i)$, $Y(T_i)$ и $A(T_i)$ для нефти марки «Brent»

2.1.2. Пример 2. Отношение цен евро к доллару

Вторым примером будет временной ряд ежедневных отношений валют евро к доллару в период с 1 ноября 2011 по 31 декабря 2013 [11]: $T_i - T_{i-1} = 1$, $i = \overline{2, n}$; $n = 543$; $A(T_i)$ — отношение евро к доллару в каждый момент времени. Выбираем сравнительно спокойный участок на фондовом рынке основных мировых валют. Определяем размах ряда $D(T_i)$ на идентификационном интервале (рис. 11). Алгоритм показал, что при $c=45$ половина размаха ряда $D(T_i)$ стала больше размаха ряда $X(T_i)$. В результате получили горизонт прогнозирования, равный 3 дням (рис. 12). При этом линейная составляющая модели $F(T_i)$ имеет вид: $Y(T_i) = -0,000081 T_i + 1,3095$.

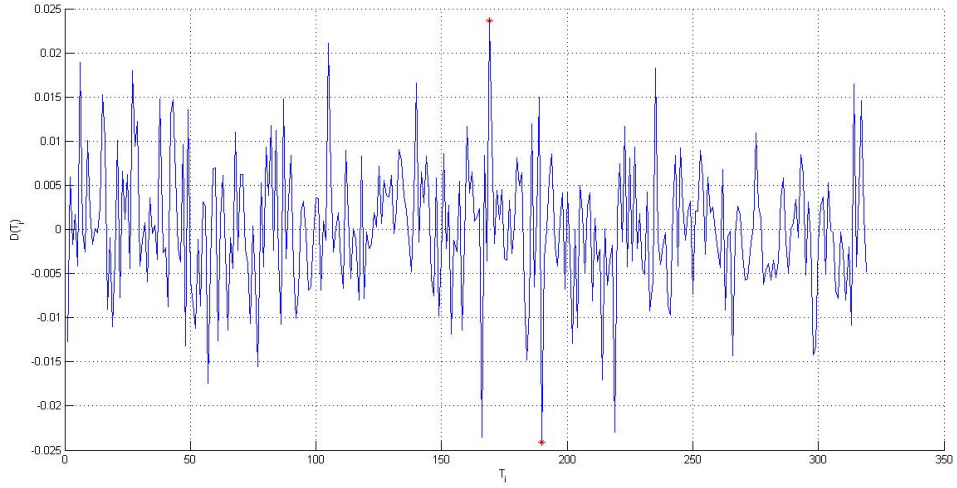


Рис. 11: $D(T_i)$ для евро/доллар

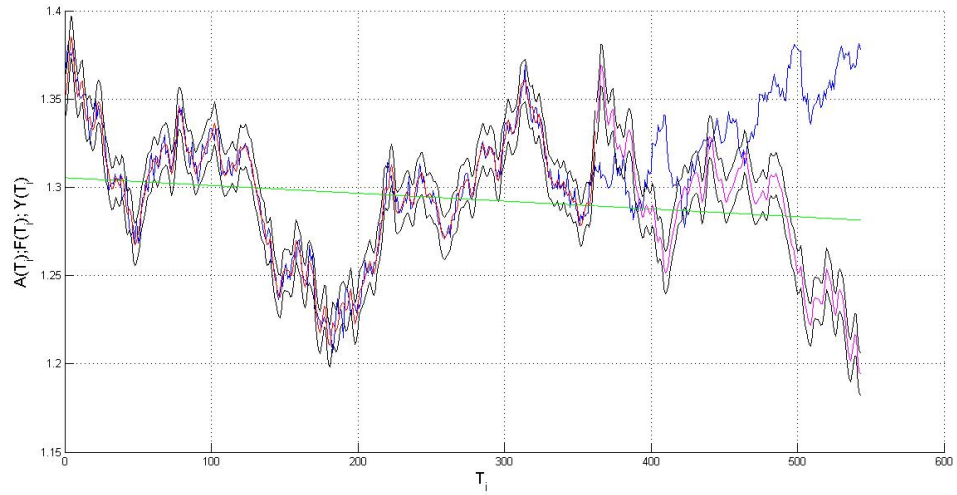


Рис. 12: $F(T_i)$, $Y(T_i)$ и $A(T_i)$ для евро/доллар

2.1.3. Пример 3. Температура в Санкт-Петербурге

Пусть дан ряд ежедневных температур в Санкт-Петербурге в период с января 2014 по сентябрь 2015 года [12]: $T_i - T_{i-1} = 1$, $i = \overline{2, n}$; $n = 607$; $A(T_i)$ — температура в каждый момент времени. Случайная составляющая в данных температурах возникает из климатических всплесков, которые обусловлены изменениями в верхних слоях атмосферы, влажности и структуры ветров в местности, которая подвергается анализу, и другим природным факторам. В совокупности эти случайные воздействия по своей абсолютной величине гораздо меньше, чем влияние Солнца и расположение Земли. Определяем размах ряда $D(T_i)$ на $2/3$ данных (рис. 13), далее действуем по алгоритму.

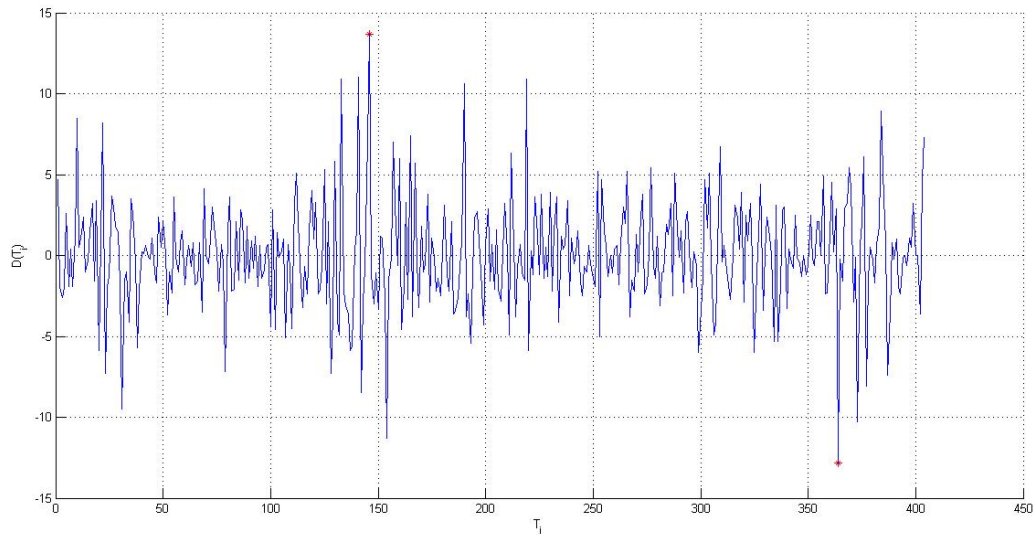


Рис. 13: $D(T_i)$ для температуры

В результате работы алгоритма получается, что половина размаха ряда $D(T_i)$ при 50 гармонике s оказалась больше размаха ряда $X(T_i)$. При этом горизонт в данном случае равен 6 (рис. 14). Линейная составляющая функции $F(T_i)$ имеет вид: $Y(T_i) = -0,0023 T_i + 7,852$. Эта добавка к данной модели весьма незначительна. Такой результат может означать, что несмотря на то, что ряд $F(T_i)$ приближает выборку, а не тренд, в данном случае $F(T_i)$ оказывается довольно хорошо приближенным и к линии, составленной из математических ожиданий, то есть случайная составляющая ряда довольно мала, либо хорошо описывается рядом $F(T_i)$.

Замечание 3. При s равном половине узлов временного ряда $A(T_i)$, $F(T_i)$ лучше всего приближает выборку $A(T_i)$. Однако нам нет надобности брать так много гармоник по двум причинам:

- 1) Необходимо приближать не реализацию случайных величин, а тренд;
- 2) При больших гармониках полоса, окружающая модельные данные, станет слишком узкой, и тогда реальные данные выйдут из этой полосы слишком рано. В этом случае маленькая величина горизонта не будет отражать суть происходящего процесса.

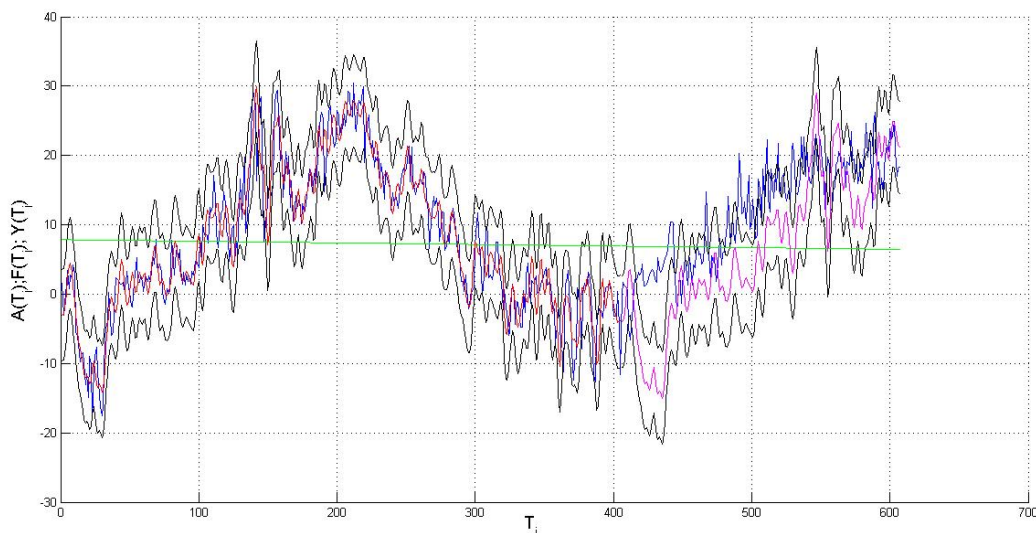


Рис. 14: $F(T_i)$, $Y(T_i)$ и $A(T_i)$ для температуры

2.2. Алгоритмы построения среднего горизонта прогнозирования

По работе алгоритма нахождения горизонта прогнозирования возникает вопрос о сохранности величины горизонта, если сместить или изменить размер идентификационного интервала. В этом параграфе будут приведены два алгоритма, решающих эту проблему, и проведён их сравнительный анализ.

Алгоритм построения среднего горизонта методом сдвига

1. Из временного ряда $A(T_i)$ выделяется его, например, первая треть.
2. На новом участке временного ряда осуществляется алгоритм построения горизонта.
3. Смещаем новый участок вправо на единицу времени. (рис. 15)
4. До тех пор, пока правая граница нового участка не достигла конца ряда $A(T_i)$, повторяем пункты 2-3.
5. Находим среднее арифметическое всех полученных горизонтов ряда и называем это значение средним горизонтом временного ряда $A(T_i)$.

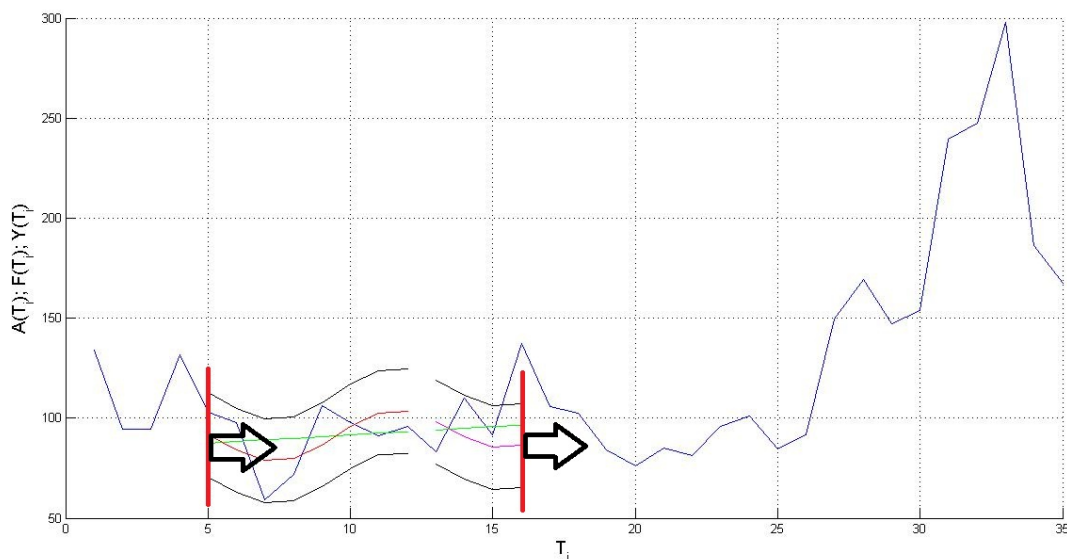


Рис. 15: Метод сдвига

Алгоритм нахождения среднего горизонта методом растяжения:

1. Рассматривается некоторая часть временного ряда, начало которого совпадает с началом ряда $A(T_i)$.
2. На новом временном ряде выполняется алгоритм нахождения горизонта.
3. Увеличиваем объём нового ряда способом увеличения значения правой границы (рис. 16).
4. Повторяем п.2-3 до тех пор, пока новый временной ряд не совпадёт с исходным.
5. Из всех полученных данных находится средний горизонт.

Был проведён сравнительный анализ двух представленных алгоритмов на примерах мировых цен на зерно [13] по годам, ежедневной температуры воздуха в Москве [14] и недельных цен на нефть марки «Brent» [15]. В результате средний горизонт был определён достаточно хорошо для обоих методов, однако алгоритм построения среднего горизонта методом растяжения имеет два существенных недостатка:

- 1) Время работы для больших временных рядов в сотни раз больше, чем в методе сдвига (так, например, средний горизонт методом растяжения для температурных данных в Москве считался порядка пяти часов, в то время как методом сдвига менее, чем за минуту). Это происходит из-за того, что при увеличении интервала идентификации увеличивается время по-

строения тренда, в то время как в методе сдвига интервал идентификации остаётся постоянным;

2) При приближении правой границы идентификационного интервала к границе исходного временного ряда $A(T_i)$ уменьшается интервал прогнозной части, то есть горизонт прогнозирования может быть меньше, чем он на самом деле является.

Таким образом, алгоритмом получения среднего горизонта будем считать метод сдвига. В результате его работы получены следующие результаты: для мировых цен на зерно средний горизонт равен 2 года; для цен на нефть марки «Brent» результат равен 1,7 недели; для температуры в Москве горизонт равен 2 дня. Такие результаты соответствуют действительности, что подтверждает правильность алгоритма.

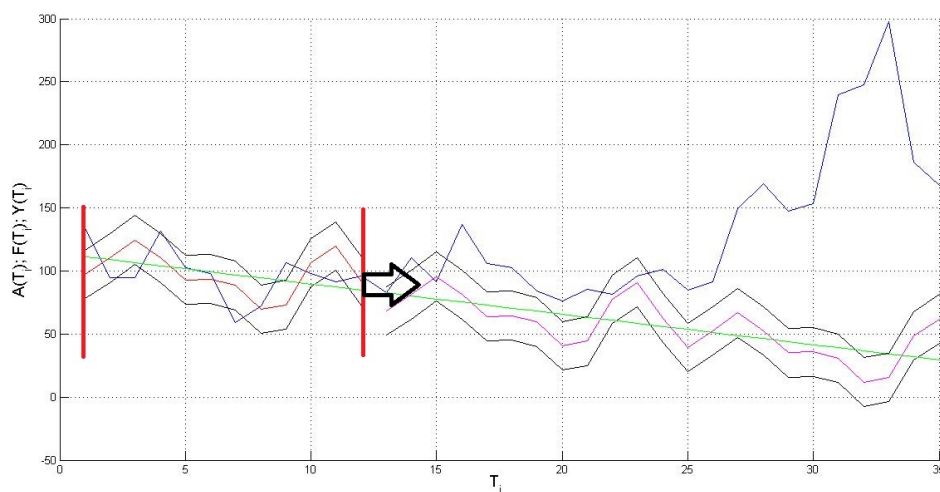


Рис. 16: Метод растяжения

Заключение

Исследованы модели временных рядов на примерах цен на нефть марки «Brent», температурных данных в Москве и Санкт-Петербурге, отношений курсов валют евро к доллару и мировых цен на зерно. В работе построена аппроксимирующая функция временного ряда с помощью выделения линейной составляющей методом наименьших квадратов и разложения в ряд Фурье. Показана целесообразность уточнения коэффициентов ряда Фурье из-за случайного характера данных. Предлагаемое уточнение ведётся по тем же самым гармоникам, что и исходный ряд.

Сформирована «полоса» вокруг модели тренда, основываясь не на отклонениях, а на «полосе» первых разностей исходного ряда, по которой определён горизонт прогнозирования. Доказано утверждение, что горизонт прогнозирования не может быть больше, чем расстояние до первого выхода реальных данных из «модельной полосы».

Приведены два алгоритма вычисления среднего горизонта временного ряда на основе ранее полученного алгоритма нахождения горизонта прогнозирования. Произведён их сравнительный анализ по времени работы. Все вычисления были выполнены в пакете MATLAB.

Список литературы

- [1] Острем К. Ю. Введение в стохастическую теорию управления. М.: Изд. Мир, 1973. 324 с.
- [2] Андерсон Т. Статистический анализ временных рядов. М.: Мир, 1976. 757 с.
- [3] Прасолов А. В., Хованов Н. В. О прогнозировании с использованием статистических и экспертных методов // Автоматика и Телемеханика. 2008., № 6. С. 129 – 143.
- [4] Prasolov A. V., Wei K. C. On Forecast of Exchange Rate of a Foreign Currency // Proc. IEEE Int. Conf. Control Appl. and IEEE Int. Sympos. Comput.-Aided Control Syst. Design, Anchorage, Alaska, USA. 2000.
- [5] Прасолов А. В. Математические методы экономической динамики: Учебное пособие. 2-е изд. испр. СПб.: Лань, 2015. 352 с.
- [6] Иванов Н. Г., Прасолов А. В. Анализ различных методов аппроксимации тренда временного ряда // Процессы управления и устойчивость. 2015. Т. 2. № 1. С. 623 – 628.
- [7] Ivanov N. G., Prasolov A. V. Analysis of the approximation of time series trend and forecasting based on it // «Stability and Control Processes» in Memory of V.I. Zubov (SCP). 2015. P. 460 – 462.
- [8] US Energy Information Administration. <http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RB RTE&f=D>
- [9] Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. 406 с
- [10] Профессиональный информационно-аналитический ресурс, посвященный интеллектуальному анализу данных. [http://www.machinelearning.ru/wiki/index.php?title=Выделение_периодической_компоненты_временного_ряда_\(пример\)](http://www.machinelearning.ru/wiki/index.php?title=Выделение_периодической_компоненты_временного_ряда_(пример))
- [11] Economic Research. Federal Reserve bank of St. Louis. <https://research.stlouisfed.org/fred2/series/DEXUSEU#>

- [12] Погода в 243 странах мира. http://rp5.ru/Weather_archive_in_Saint_Petersburg
- [13] Economic Research. Federal Reserve bank of St. Louis. <https://research.stlouisfed.org/fred2/series/PMAIZMTUSDM>
- [14] Average Daily Temperature Archive. <http://academic.udayton.edu/kissock/http/Weather/gsod95-current/RSMOSCOW.txt>.
- [15] Economic Research. Federal Reserve bank of St. Louis. <https://research.stlouisfed.org/fred2/series/POILBREUSDM>

Приложение

Программная реализация в пакете MATLAB

Построение детерминированной части тренда.

```
A = load('brent.txt');
n = size(A,1);
A=A(1:n,1);
T = row(A,n);
%p отвечает за то, чтобы замеры начинались сначала
%(нужно для прогнозирования)
p=1;
[y,k] = regress1(T,A,n);
[ys,R] = differ1(n,k,T,A,y);
if rem(n,2)==0
    c = n/2;
else
    c = (n-1)/2;
end
s=c;
%вычислим значения  $X(t)=R(t)-Y(t)$  в каждой точке временного ряда
[a,b,X]=numFour(p,n,c,R,T);
[X,Pogr,ak,bk] = mnkFour(p,n,c,R,T);
figure
[D,S]=dispR_F(p,s,n,T,R);
E=dispR_F_Pogr(s,n,T,R,p);

%построение временного ряда по массиву
function T = row(A,n)
T=zeros(n,1);
    for i=1:n
        T(i,1)=i;
    end
figure
hold on
```

```

grid on
plot(T,A)
title('Нефть марки "Brent"');
xlabel('Время (нед.)');
ylabel('Цена за барель ($)');
end

%выделение линейной составляющей
function [y,k]=regress1(T,A,n)
    k=polyfit(T,A,1);
    syms t
    y=k(1)*t+k(2);
end

%построение ряда R(T_i)
function [ys,R] = differ1(n,k,T,A,y)
    ys=zeros(n,1);
    R=zeros(n,1);
    for i=1:n
        ys(i,1)=k(1)*T(i,1)+k(2);
        R(i,1)=A(i,1)-ys(i,1);
    end
end

%построение ряда Фурье
function [a,b,X,F]=numFour(p,n,c,R,T)
    F=zeros(n,1);
    a=zeros(c,1);
    b=zeros(c,1);
    for t=1:n
        sumF=0;
        for k=1:c
            sumRa=0;
            sumRb=0;
            for i=1:n

```

```

        sumRa=sumRa+R(T(i,1)-p+1,1)*cos(2*pi*k/n*(T(i,1)-p+1));
        sumRb=sumRb+R(T(i,1)-p+1,1)*sin(2*pi*k/n*(T(i,1)-p+1));
    end
    a(k,1)=2/n*sumRa;
    b(k,1)=2/n*sumRb;
    sumF=sumF+a(k)*cos(2*pi*k/n*t)+b(k)*sin(2*pi*k/n*t);
    end
    F(t,1)=sumF;
end
X=R-F;
end

```

%оптимизация коэффициентов Фурье

```
function [X,Pogr,ak,bk] = mnkFour(p,n,c,R,T)
```

```

m=2*c;
Q=zeros(n,m);
ak=zeros(c,1);
bk=zeros(c,1);
for i=1:n
    for j=1:2:2*c
        Q(i,j)=cos(2*pi*j/n*T(i,1));
        Q(i,j+1)=sin(2*pi*j/n*T(i,1));
    end
end
H=Q'*Q;
[a,b,X]=numFour(p,n,c,R,T);
B=Q'*X;
kf=H\B;
for i=1:c
    ak(i,1)=kf(2*i-1)+a(i,1);
Pogr=zeros(n,1);
for i=1:n
    for k=1:2:2*c
        Pogr(i,1)=Pogr(i,1)+kf(k)*cos(2*pi*k/n*T(i,1))+...
        ...kf(k+1)*sin(2*pi*k/n*T(i,1));
    end
end

```

```

        end
    end
end

%Построение графиков квадрата ошибки отклонения D(c) и E(c)
function [D,S]=dispR_F(p,s,n,T,R)
    S=zeros(s,1);
    D=zeros(s,1);
    for c=1 : s
        c
        S(c,1)=c;
        [a,b,X]=numFour(p,n,c,R,T);
        D(c,1)=(norm(X))^2;
    end
    figure
    plot(S,D);
    hold on
    grid on
    title(' ');
    xlabel('c');
    ylabel('D(c)');
end

function E=dispR_F_Pogr(s,n,T,R,p)
    E=zeros(s,1);
    S=zeros(s,1);
    for c=1 : s
        S(c,1)=c;
        [X,Pogr] = mnkFour(p,n,c,R,T);
        E(c,1)=vpa((norm(X-Pogr))^2);
    end
    hold on
    plot(S,E,'r');
    grid on
    title(' ');

```

```

xlabel('c');
ylabel('D(c),E(c)');
end

```

Алгоритм построения горизонта прогнозирования

```

A = load('brent.txt');
n = size(A,1)
A=A(1:n,1);
T = row(A,n);
p=1;
[y,k] = regress1(T,A,n);
[ys,R] = differ1(n,k,T,A,y);
if rem(n,2)==0
    c = n/2;
else
    c = (n-1)/2;
end
s=c;
[razmah,D]=first_razn_row(A,T,n)
horiz = connect_len(razmah,c,n,T,R,ys,A,p);

%построение ряда первых разностей
function [razmah,D]=first_razn_row(R,T,n)
n1=n-1;
D=zeros(n1,1);
T1=T(1:n1,1);
for i=1:n1
    D(i,1)=R(i,1)-R(i+1,1);
end
[M,I]=max(D);
[m,i]=min(D);
razmah=M-m
stripwidth=razmah/2
figure
hold on

```

```

grid on
plot(T1,D,'b*-')
plot(i,m,'r*')
plot(I,M,'r*')
xlabel('T_i')
ylabel('D(T_i)')
end

```

%Функция, выполняющая определение горизонта ряда и оптимальное
%количество гармоник 'с'

```

function horiz = connect_len(razmah,c,n,T,R,ys,A,p)
n1=round(2/3*n);
T1 = T(1:n1,1);
A1=A(1:n1,1);
[y,k] = regress1(T1,A1,n1);
vpa(y,12)
[ys1,R1] = differ1(n1,k,T1,A1,y);
T2=T(n1+1:n,1);
A2=A(n1+1:n,1);
[ys2,R2] = differ1(n-n1,k,T2,A2,y);
c1=c;
c=0;
flag=false;
while flag==false && c<c1
    c=c+1;
    [X1,Pogr1,len]=gist_len(c,n1,T1,R1,p);
    if len<razmah/2
        flag=true;
    end
end
F1=R1-X1-Pogr1;
F1_pr=zeros(n-n1,1);
flag1=false;
for i=1:n-n1
    %копируем последовательно данные из R1 в R2

```

```

    if rem(i,n1)==0
        F1_pr(i,1)=F1(i,1);
    else
        F1_pr(i,1)=F1(rem(i,n1),1);
    end
    %определяем выход модельных данных из модельной полосы
    if flag1==false
        if F1_pr(i,1)+razmah/4+ys2(i,1) < R2(i,1)+ys2(i,1)...
            ...|| F1_pr(i,1)-razmah/4+ys2(i,1) > R2(i,1)+ys2(i,1)
                flag1=true;
                horiz=i;
                {'c=' c 'gorizont=' i}
            end
        end
    end
    if flag1==false
        flag1=true
        horiz=i
        {'c=' c 'gorizont=' i}
    end
end
end

```

Алгоритм построения среднего горизонта методом сдвига

```

A = load('brent.txt');
n = size(A,1);
A=A(1:n,1);
T = row(A,n);
p=1;
[y,k] = regress1(T,A,n)
[ys,R] = differ1(n,k,T,A,y);
n1=round(1/3*n); %назначаем величину скользящего ряда (вместе с
% прогнозным интервалом)
n2=round(2/3*n1); %скользящая величина глубины прогноза
if rem(n2,2)==0
    c = n2/2;

```



```

else
    c = (n2-1)/2;
end
s=0;
for p=1:n-n1+1
    p
    razmah=first_razn_row_sdvig(R,T,n1,p);
    [harm, horiz] = connect_len_sdvig(razmah,c,n,T,R,ys,A,n1,n2,p);
    %gor=mean_exits(c,i,n1,n2,A,T);
    if horiz==n1-n2
        s=s+1;
    end
    H(p,1)=horiz;
    C(p,1)=harm;
    D(p,1)=razmah/2;
end
vpa([H C D],6)
[s s/size(H,1)*100]
{'ср. горизонт=' sum(H)/(n-n1+1) 'ср. число гармоник=' ...
...sum(C)/(n-n1+1) 'ср. ширина полосы=' sum(D)/(n-n1+1)}

```

%ряд первых разностей для текущего идентификационного интервала

```
function [razmah,D]=first_razn_row_sdvig(R,T,n1,p)
```

```
D=zeros(n1-1,1);
```

```
T1=T(p:p+n1-1,1);
```

```
for i=p:p+n1-2
```

```
    D(i-p+1,1)=R(i,1)-R(i+1,1);
```

```
end
```

```
[M,I]=max(D);
```

```
[m,i]=min(D);
```

```
razmah=M-m;
```

```
end
```

%Функция, выполняющая определение горизонта ряда и оптимальное
%количество гармоник 'с' для текущего идентификационного интервала

```

function [harm, horiz] = ...
...connect_len_sdvig (razmah, c, n, T, R, ys, A, n1, n2, i)
%строим временной ряд - ряд глубины прогноза
w=i;
T1 = T(i:i+n2-1,1);
A1 = A(i:i+n2-1,1);
[y,k] = regress1(T1,A1,n2);
[ys1,R1] = differ1(n2,k,T1,A1,y);

%строим ряд для определения прогноза
T2 = T(i+n2:i+n1-1,1);
A2 = A(i+n2:i+n1-1,1);
[ys2,R2] = differ1(n1-n2,k,T2,A2,y);

c1 = c;
c = 0;
flag=false;
while flag==false && c<c1
    c=c+1;
    [X1,Pogr1,len]=gist_len(c,n2,T1,R1,i);
    if len<razmah/2
        flag=true;
    end
end
F1=R1-X1-Pogr1;
F1_pr=zeros(n1-n2,1);
flag1=false;
for i=1:n1-n2
    %копируем последовательно данные из R1 в R2
    if rem(i,n2)==0
        F1_pr(i,1)=F1(n2,1);
    else
        F1_pr(i,1)=F1(rem(i,n2),1);
    end
end
%определяем выход модельных данных из трубы

```

```

if flag1==false
    if F1_pr(i,1)+razmah/4+ys2(i,1) < R2(i,1)+ys2(i,1)...
        ... || F1_pr(i,1)-razmah/4+ys2(i,1) > R2(i,1)+ys2(i,1)
            flag1=true;
            horiz=i;
            { 'gorizont=' i 'с=' с 'ширина полосы=' razmah/2}
            harm=c;
        end
    end
end
if flag1==false
    flag1=true;
    horiz=i;
    {'с=' с 'gorizont=' i 'ширина полосы=' razmah/2}
    harm=c;
end
end

```